NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

ANALYSIS OF REGIONAL EFFECTS ON MARKET SEGMENT PRODUCTION

by

James D. Moffitt

June 2016

Thesis Advisor: Lyn R. Whitaker
Co-Advisor: Jonathan K. Alt
Second Reader: Jeffrey B. House

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704–0188*

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE June 2016	3. REPORT TYPE AND DATES COVERED Master's thesis

4. TITLE AND SUBTITLE
ANALYSIS OF REGIONAL EFFECTS ON MARKET SEGMENT PRODUCTION

5. FUNDING NUMBERS

6. AUTHOR(S) James D. Moffitt

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES)
United States Army Recruiting Command
Fort Knox, KY

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number ____N/A____.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (maximum 200 words)

This thesis develops a data-driven statistical model capable of identifying regional factors that affect the number of United States Army Recruiting Command (USAREC) accessions in Potential Rating Index Zip Code Market New Evolution (PRIZM NE) market segments. This model will aid USAREC G2 analysts involved in conducting recruiting market intelligence. Market intelligence helps the commander visualize the performance of subordinate units within their market and provides recommendations for use and expansion. This thesis first attempts to establish that a single high-assessing PRIZM NE market segment, Segment 32, does not access recruits at the same rate across regions. This thesis then develops general linear regression and gradient boosted decision tree models to determine the regional factors that contribute to the variance of recruit production. In particular, the gradient boosted decision tree delivers predictive results that allow analysts to identify regions that have underperforming accession rates compared to the national average. The recommendation of this thesis is that the USAREC implement the gradient boosted decision trees for use in G2 market analysis.

14. SUBJECT TERMS
recruiting, market segmentation, PRIZM NE, Poisson regression, gradient boosted decision tree

15. NUMBER OF PAGES
79

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UU

NSN 7540–01-280-5500

Standard Form 298 (Rev. 2–89)
Prescribed by ANSI Std. 239–18

THIS PAGE INTENTIONALLY LEFT BLANK

ANALYSIS OF REGIONAL EFFECTS ON MARKET SEGMENT PRODUCTION

James D. Moffitt
Captain, United States Army
B.S., Texas A&M University, 2005

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
June 2016

Approved by: Lyn R. Whitaker
Thesis Advisor

Jonathan K. Alt
Co-Advisor

Jeffrey B. House
Second Reader

Patricia A. Jacobs
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This thesis develops a data-driven statistical model capable of identifying regional factors that affect the number of United States Army Recruiting Command (USAREC) accessions in Potential Rating Index Zip Code Market New Evolution (PRIZM NE) market segments. This model will aid USAREC G2 analysts involved in conducting recruiting market intelligence. Market intelligence helps the commander visualize the performance of subordinate units within their market and provides recommendations for use and expansion. This thesis first attempts to establish that a single high-assessing PRIZM NE market segment, Segment 32, does not access recruits at the same rate across regions. This thesis then develops general linear regression and gradient boosted decision tree models to determine the regional factors that contribute to the variance of recruit production. In particular, the gradient-boosted decision tree delivers predictive results that allow analysts to identify regions that have underperforming accession rates compared to the national average. The recommendation of this thesis is that the USAREC implement the gradient boosted decision trees for use in G2 market analysis.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

AC	Active Component
ACS	Army Custom Segments
AGI	Adjusted Gross Income
AOR	Area of Responsibility
AUC	Area Under the Curve
BDE	Brigade
BN	Battalion
CB	Census Bureau
CBSA	Core Based Statistical Area
CDC	Center for Disease Control
CHSI	Community Health Status Indicators
DOD	Department of Defense
FBI	Federal Bureau of Investigation
FY	Fiscal year
GLM	Generalized Linear Model
HUD	Department of Housing and Urban Development
IPEDS	Integrated Postsecondary Education Data System
JAMRS	Joint Advertising Market Research and Studies
MLR	Multiple Linear Regression
MRB	Medical Recruiting Brigade
PRIZM NE	Potential Rating Index Zip Code Market New Evolution
QMA	Qualified Military Available
ROC	Receiver Operating Characteristic
SAMA	Segmentation Analysis and Market Assessment
SE	Standard Error
USAREC	United States Army Recruiting Command
USPS	United States Postal Service
ZCTA	ZIP Code Tabulation Area
ZIP	Zone Improvement Plan

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

United States Army Recruiting Command (USAREC) provides the manning for the active Army and the U.S. Army Reserve, ensuring security and readiness for our Nation (USAREC, 2009). USAREC contains approximately 9,500 personnel, 6 recruiting brigades, 38 battalions, 258 companies, and over 1,400 recruiting stations (USAREC, 2015). Of the roughly 34 million youths eligible for military service, United States Army recruiters must find 62,000 recruits to satisfy Congressionally mandated end-strength goals (Feeney, 2014; Tice, 2016). In order to meet those goals, USAREC must understand the regional socio-economic factors that affect recruiting. Knowledge of factors that affect recruiting allows USAREC to more efficiently allocate resources to maximize recruit production.

Currently, USAREC uses a segmentation analysis and market assessment tool (SAMA) to identify, prioritize, and target recruiting markets in recruiting station areas of operation (Clingan, 2011). SAMA uses Nielsen Potential Rating Index for ZIP Markets New Evolution (PRIZM NE) data to generate ZIP code-based reports that show recruiters which ZIP codes are “must keep,” “must win,” and which are “markets of opportunity.” SAMA also generates real-time production data by PRIZM NE market segment as well as historical production. The data generated by SAMA is used to access recruiting center potential (Marmion, 2015).

This research collects data sets similar to those identified in previous research as relevant for predicting the number of recruiting accessions (Intrater, 2015). The data sets are open source and include economic and demographic data from the Census Bureau’s American Community Survey, health data from the Center for Disease Control’s Community Health Status Indicators, and crime data from the Department of Housing and Urban Development. Data is collected at the ZIP code tabulation area level and the county level, and then aggregated to Core Based Statistical Areas (CBSA). This research uses CBSA to define regions. A CBSA consists of an area with at least one core population nucleus of at least 10,000 plus adjacent areas having a high degree of social and economic integration with the core (United States Census Bureau, 2013).

We develop Poisson generalized linear models (GLM) and gradient boosted decision trees to help identify the regional characteristics that affect PRIZM NE market segment production across CBSAs. We first establish that market segments have different accession rates across the country. Next we analyze a top-accessing PRIZM NE market segment, Segment 32, to develop the methodology to find regional characteristics. Once the methodology is established, we apply it to the top five accessing PRIZM NE market segments.

The research finds that both Poisson GLMs and gradient boosted decision trees can be used to determine region factors that affect PRIZM NE market segment recruit production, and also that gradient boosted decision trees can predict a CBSA's ability to under- or over-perform compared to the mean PRIZM NE market segment accession rate. These models provide USAREC with a methodology to access market segment accession rate variance when they did not have a methodology before.

References

Clingan, L. (2011). SAMA not just another acronym. *Recruiter Journal*, *63*(6), 27.

Feeney, N. (2014, June 29). Pentagon: 7 in10 youths would fail to qualify for military service. *Time*. Retrieved from http://time.com/2938158/youth-fail-to-qualify-military-service/

Intrater, B. C. (2015). Understanding the impact of socio-economic factors on Navy accessions (Master's thesis). Naval Postgraduate School. Retrieved from Calhoun http://hdl.handle.net/10945/47279

Tice, J. (2016, February 23). Army recruiting market tightens but service expects to make 2016 goal. *Army Times*. Retrieved from http://www.armytimes.com/story/military/careers/army/2016/02/23/army-recruiting-market-tightens-but-service-expects-make-2016-goal/80624982/

United States Census Bureau. (2013). Maps of metropolitan and micropolitan statistical areas. Retrieved from http://www.census.gov/population/metro

USAREC. (2009). *Recruiting operations*. Fort Knox, KY: Headquarters, USAREC.

USAREC. (2015). USAREC—About us. Retrieved from USAREC: http://www.usarec.army.mil/aboutus.html

ACKNOWLEDGMENTS

I would like to recognize and thank all of the people who supported me throughout this thesis process. Your encouragement, knowledge, and selfless assistance were invaluable to the completion of my thesis.

First, I would like to thank my wife, Alyssa, for her endless support, encouragement, understanding, and love throughout my career and time here at NPS. She sets the standard for hard work and selfless service that I try to follow every day. Thank you to my father and mother for instilling in me a hard work ethic and the will to never give up.

Next, I would like to thank the NPS manpower-working group, Professor Lyn Whitaker, Professor Sam Buttrey, LTC Jonathan Alt, LTC Jeffrey House, Major Brandon Fulton, Major Glenn Darrow, and Captain Emilie Monaghan. The encouragement, sharing of ideas, and guidance significantly contributed to my thesis and Army and Navy recruiting.

Finally, I would like to thank the USAREC G2 team (Michael Nelson, Mitch Stokan, Joe Baird, and Major David Devin) for providing great real-world problems for us to solve and for their excellent facilitation and continuous support.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

Of the roughly 34 million youths eligible for military service, United States Army recruiters must find 62,000 recruits to satisfy Congressionally mandated end-strength goals (Feeney 2014; Tice 2016). The United States Army Recruiting Command (USAREC) must understand the socio-economic terrain of a recruiting area to identify and overcome barriers to recruiting to efficiently allocate resources to maximize recruiting production.

Widely used by civilian industry, Potential Rating Index for ZIP Code New Evolution (PRIZM NE) market segments provide users a better understanding of their customers (Nielsen, 2016). Nielsen assigns every household in America to a PRIZM NE market segment based on demographics, credit card spending, media usage, and leisure activities. Nielsen then aggregates the data to the Zone Improvement Plan (ZIP) code +4 level, which is typically an area consisting of 10 to 12 households (Nielsen, 2016). USAREC uses PRIZM NE market segments to build models that access recruiting center market potential (Marmion, 2015). In the building of these models USAREC operates as if accession rate is constant over regions for each of the 66 PRIZM NE market segments.

The research in this thesis identifies factors affecting Army, active component (AC), accessions; explores the development of statistical models with socio-economic factors; and uses those models to help identify if there are regional differences in PRIZM NE market segments. This research provides USAREC's G2 (Market Analysis Division) models at the regional level, defined as a Core Based Statistical Area (CBSA), using open-source ZIP code-level data.

A. PROBLEM STATEMENT

A strong economy, reduced unemployment rates, and increased obesity rates limit the population of available people qualified for military service by a third creating a significant challenge for Army recruiters to meet missioning goals (McHugh & Odierno, 2015). This research develops statistical models to better understand the influence of regional effects on the number of accessions in a geographic area. We address this problem by answering these contributing questions:

- Do PRIZM NE market segments produce the same number of recruits per capita regardless of region?
- Within a PRIZM NE market segment which population factors affect recruit production?

The outcome of this research provides USAREC leadership with the ability to identify regional characteristics that make it harder to recruit in certain markets. The insights gained will better inform decisions related to recruiter placement, and recruiting station realignment to meet annual recruiting goals.

B. RESEARCH OVERVIEW

We divide this report into five chapters. Chapter II covers USAREC background, market segmentation, and a literature review of past work related to this thesis. Chapter III covers the data and constraints, assumptions, and limitations for the models. Chapter IV focuses on the model development, output, and analysis. Chapter V outlines recommendations and discusses future work.

II. BACKGROUND AND LITERATURE REVIEW

Chapter II provides a brief overview of USAREC operations, USAREC G2 Market Analysis, and the market segmentation tools G2 uses to analyze market production. Finally, we review four academic studies that have leveraged socio-economic data to identify predictors and estimate market potential.

A. UNITED STATES ARMY RECRUITING COMMAND

USAREC provides the manning for the Active Army and the U.S. Army Reserve, ensuring security and readiness for our Nation (USAREC, 2009). USAREC contains approximately 9,500 personnel, six recruiting brigades, 38 battalions, 258 companies, and over 1,400 recruiting stations (USAREC, 2015). Figure 1 shows the area of responsibility (AOR) for the five recruiting brigades (BDE), the medical recruiting brigade (MRB), and their subordinate battalions (BN). Each color represents a different brigade and the black lines depict battalion boundaries.

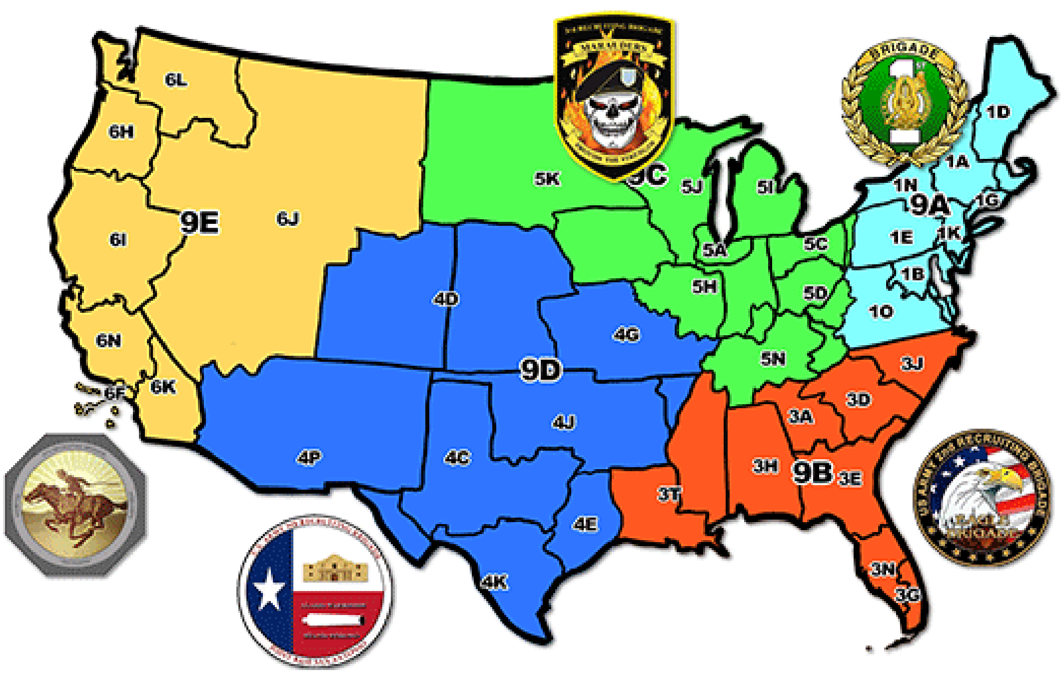


Figure 1. United States Army Recruiting Command Brigade and Battalion Boundaries. Source: USAREC (2015).

B. USAREC G2

USAREC G2 conducts market intelligence, mission analysis, and the missioning process for the USAREC Commander (USAREC, 2009). Market intelligence helps the commander visualize the performance of subordinate units within their market and provides recommendations for use and expansion. G2 develops mission analysis tools to assist in the allocation of recruiters to maximize recruiting yield. In the missioning process G2 distributes the annual accessions mission based on recruiting area market potential (USAREC, 2009). USAREC uses the number of Qualified Military Available and a four-year weighted average of past recruitment production to access market depth and assign recruiting missions (Marmion, 2015).

C. MARKET SEGMENTATION AND CURRENT PRACTICES

Since 2005, the Army has used the Army Custom Segments (ACS) created under contract by Integras to partition the United States youth population to better understand the recruiting market. To construct the ACS, Integras conducted surveys of the youth population to determine motivators and barriers to enlistment. They then combine the survey results with factor analysis of the Nielsen company's 66 Potential Rating Index for ZIP Code New Evolution (PRIZM NE) market segments to combine the segments into 39 Army Custom Segments (Devin, personal communication, March 24, 2016). USAREC and Marmion find that ACSs hinder lead development and the assessment of market potential. As a result, USAREC now uses PRIZM NE market segments to better access market potential and to develop leads (Marmion, 2015).

USAREC uses the Segmentation Analysis and Market Assessment tool (SAMA) to identify, prioritize, and target all markets in an area of operation (Clingan, 2011). SAMA uses Nielsen PRIZM NE data to generate ZIP code-based reports that show recruiters which Zone Improvement Plan (ZIP) codes are "must keep," "must win," and which are "markets of opportunity." SAMA also generates real-time production data by PRIZM NE market segment as well as historical production. The data generated by SAMA is used to access recruiting center potential.

D. GEOGRAPHIC CLASSIFICATIONS

1. ZIP Codes

The United States Postal Service (USPS) defines ZIP code boundaries and changes them annually based on the availability of postal services (United States Postal Service, 2016). ZIP code types include unique high volume addresses, post office box only, military, and standard addresses the USPS designates ZIP code boundaries based on proximity to the nearest post office (United States Postal Service Office of the Inspector General, 2013). The Census Bureau and other agencies interpolate ZIP code areas using an annually generated key to create maps and shape files. We find that ZIP codes are a poor measure for researchers because they do not represent real geographic areas and change often.

2. ZIP Code Tabulation Area

The Census Bureau created ZIP Code Tabulation Areas (ZCTA), to represent the physical spaces of the United States Postal Service ZIP Codes, to overcome the data collection problems inherent in ZIP Codes (United States Census Bureau, 2016b). The Census Bureau assigns blocks that do not have ZIP Codes to ZCTAs based on shared boundaries. From this research, we conclude that the ZCTA possesses greater utility than ZIP codes because the Census Bureau collects most of its data at this level and we can easily aggregate it.

3. Core Based Statistical Areas

A Core Based Statistical Area (CBSA) consists of an area with a population nucleus of at least 10,000 plus adjacent areas having a high degree of social and economic integration with the core (United States Census Bureau, 2013). The Office of Management and Budget divides CBSAs by population into metropolitan or micropolitan. Metropolitan CBSAs possess population greater than or equal to 50,000, while micropolitan CBSAs have populations greater than 10,000 but less than 50,000 (United States Office of Management and Budget, 2015). Figure 2 shows the United

States CBSA boundaries; metropolitan CBSAs are dark green, micropolitan CBSAs are light green, and areas without CBSAs are white.

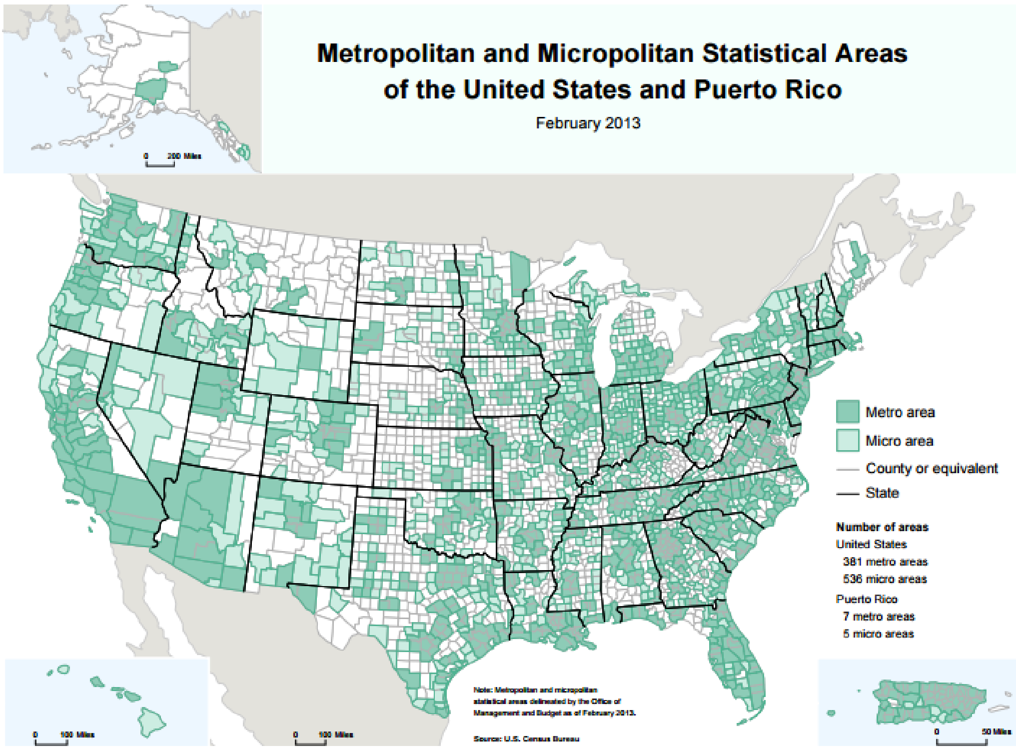


Figure 2. U.S. Metropolitan and Micropolitan Statistical Areas.
Source: United States Census Bureau (2013).

E. RELATED WORKS

Marmion (2015) finds that by using the Army Custom Segments (ACS), Segmentation Analysis and Market Assessment tool (SAMA) over-predicts the number of potential accessions by more than 25 percent for 96 percent of the centers and that the average prediction is 35 percent over the 2014 production. Marmion finds that when the SAMA model uses PRIZM NE market segments the model over-predicts potential by 41 percent. To develop a better predictive tool, Marmion (2015) first creates two new factors from the PRIZM NE data: a core score similar to a market penetration rate and a social score based on urbanization. He then creates a multiple linear regression model (MLR) that incorporates the two PRIZM factors, a brigade identifier, and four-year weight average of contracts that produces an R-squared value better than the current implementation of SAMA.

An important consideration when attempting to predict recruiting potential is the socio-economic conditions in an area. Jackson finds that the factors with the most predictive power include the number of recruiters, QMA (Qualified Military Available), and unemployment rates (Jackson, 2015). Jackson’s work serves as a baseline for which factors we include our recruiting models. We develop a model that does not include the recruiters as a factor.

In addition to socio-economic conditions, other characteristics influence an area’s potential for recruiting. Intrater finds that the most influential factors affecting Navy accessions in the open source data were the number of recruiters assigned, adjusted gross income (AGI), the veteran population, and the number of universities in the area (Intrater, 2015). Factors that impact Navy recruiting may not correlate with facts that impact Army recruiting. This research builds on Intrater’s analysis and incorporates PRIZM NE data to determine regional effects on Army accession production.

F. SCOPE OF RESEARCH

This research uses statistical models to identify which independent variables are associated with the variability in recruit production across CBSAs. We use economic, demographic, health, voting, and education data to construct dependent variables. All data are from 2010 to present. We focus only on United States CBSAs that occur within the United States and exclude U.S. territories.

1. Constraints

USAREC requests that the analysis examine the performance of PRIZM NE Segment 47 across CBSAs.

2. Limitations

We limit data used for this research to data already available to the sponsor and publically available data sets. This allows the sponsor to implement the findings of this research and to replicate the same methods as new data become available.

We also limit the scope of this research to CBSAs inside the United States, omitting data from Puerto Rico and other U.S. territories. This limitation is necessary because data for U.S. territories is incomplete.

Additionally, this research will only focus on regular Army enlisted soldier recruitment. USAREC is also responsible for recruiting officers and enlisted members for the Army Reserve. Parker (2015) finds that the main determinant of a reserve unit's ability to meet manning requirements is strongly tied to reserve basing so reserve recruiting is outside the scope of our research.

The final limitation is the availability, completeness, and timing of the data used in this research. Government agencies do not typically collect data at the CBSA level; data is normally collected at the county level and below. We will aggregate data to the CBSA level using methods described in Chapter III. There is variability in the time periods in which some of the data is available; for instance, Census data is only collected every 10 years, because of that we use data from the Census Bureau's American Community Survey that approximates annual data from smaller scale annual surveys. See Chapter III for a more detailed discussion of the data used.

3. Assumptions

Due to the limitations of our data, we aggregate all county-level data to CBSA-level using relationship files from the U.S. Census Bureau and the Department of Housing and Urban Development (HUD). Where boundaries do not match, the relationship files partition the lower-level unit into the higher-level unit based on the percentage of residential addresses present or the percentage of the land present. We assume the distribution of population throughout counties, ZIP codes, and ZIP code tabulation areas is homogenous. We acknowledge that this is most likely not true, but this is the best method we have available.

III. DATA COLLECTION AND METHODOLOGY

In Chapter III, we discuss the data gathering and preparation required for model building. USAREC and government websites—such as the Census Bureau (CB) and Federal Bureau of Investigations (FBI)—provide data at ZIP Code Tabulation Area (ZCTA), ZIP code, and county levels. USAREC G2 provides contract performance and population data by ZIP code.

A. DATA

1. USAREC PRIZM NE Data Set

The data set provided by USAREC G2 includes the number of enlisted accessions at the ZIP code level between fiscal year 2011 (FY) and FY2014 (Devin, personal communication, March 24, 2016). For each ZIP Code, the data gives the Potential Rating Index Zip Code Market New Evolution (PRIZM NE) market segment accession and total population counts. We convert the data set from ZIP code level to CBSA level with the aid of a ZIP to CBSA crosswalk from the Department of Housing and Urban Development (HUD) (see Appendix A).

The data contains the number of accessions not associated with a PRIZM NE market segment or a CBSA. We exclude accessions observations with no PRIZM NE market segment or no CBSA from the dataset; see Table 1 for more details on the excluded data.

Table 1. Uncategorized Accession Observations.

Year	Total Accessions	Accessions Missing PRIZM Segment	% Missing PRIZM Segment	Accessions Missing CBSA	% Missing CBSA	% Missing CBSA or Segment
2011	72619	5450	7.5	4586	6.3	12.5
2012	61795	3598	5.8	3779	6.1	10.4
2013	65894	3768	5.7	4056	6.2	10.4
2014	59029	4241	7.2	3506	5.9	11.5

Joint Advertising Market Research Studies (JAMRS) define high accessing market segments as segments that account for at least two percent of total accessions counts and have an index score of 115 or above. The JAMRS index score is a priority score that measures a market segment's media consumption relative to the national consumption (Joint Advertising Market Research & Studies, 2014). Table 2 shows the top thirteen accessing market segments according to JAMRS. We find from the USAREC PRIZM NE data that the JAMRS top accessing PRIZM NE market segments are not always the top accessing market segments. Each year, 2011 to 2014, three segments are in the top 13 accessing market segments but not listed within the JAMRS top 13. This comparison only takes into account accessions and does not factor in an index score. The difference between the data's top 13 segments and JAMRS top 13 market segments is a decrease of 700–1000 recruit candidates annually. We approximate the JAMRS index score by instead using the market segment penetration rate, defined as (Marmion, 2015).

$$\text{Market Segment Penetration Rate } = \frac{\left(\frac{Accessions\ in\ Segment}{Total\ Accessions\ of\ all\ segments}\right)}{\left(\frac{Population\ in\ Segment}{Total\ Population\ of\ all\ segments}\right)} \qquad (3.1)$$

When we filter the top 13 accessing market segments by penetration rate, the difference in market segments between the actual top 13 and JAMRS top 13 drops from 3 to 2 segments. By increasing the penetration rate threshold to greater than or equal to 1.15, the actual top 13 market segments mirror the JAMRS top 13 market segments.

Table 3 shows a more detailed picture of the difference between the top 13 accessing market segments this research found and the top 13 market segments proposed by JAMRS. Limiting recruiting efforts to only the top JAMRS segments could lead to a loss of roughly 1,000 recruit candidates per year.

Table 2. JAMRS Top Accessing PRIZM NE Market Segments for U.S. Army. Adapted from Joint Advertising Market Research & Studies (2014).

Segment	Segment Name
13	Upward Bound
18	Kids & Cul-de-Sacs
20	Fast-Track Families
32	New Homesteaders
33	Big Sky Families
34	White Picket Fences
36	Blue-Chip Blues
37	Mayberry-ville
41	Sunset City Blues
45	Blue Highways
50	Kid Country, USA
51	Shotguns & Pickups

Table 3. The Data's Top 13 Accessing Market Segments vs. the Top 13 Market Segments Proposed by JAMRS. Adapted from: Joint Advertising Market Research & Studies (2014).

Year	Data Top 13 Segments Accessions	JAMRS Top 13 Segment Accessions	Difference	Top Segments Not in JAMRS	In JAMRS Not in Top Data	Top Segments Not in JAMRS with a PenRate > 1.0	Top Segments Not in JAMRS with a PenRate > 1.15
2011	23,102	22,011	1,091	48, 63, 64	36, 41, 45	43 ,64	19, 43
2012	20,120	19,249	871	48, 63, 64	36, 41, 45	43, 64	43, 64
2013	21,199	20,422	777	29, 63, 64	36, 45, 50	63, 64	None
2014	18,746	17,885	861	29, 63, 64	36, 41, 45	63, 64	None

The American Community Survey provides annual information about housing, education, employment, and veterans (United States Census Bureau, 2016a). The Census Bureau provides data in 1-year, 3-year, and 5-year average data profiles. We use the 5-year average data profiles in this research because the average takes more data into account.

The American Community Survey includes a large number of reports at varying levels of detail. We selected forty-nine key variables from those reports for this research based on previous research (Intrater, 2015). We use variables such as educational attainment, employment status, veteran population, household income, and workers by industry from this data set. For a more detailed account of variable selection, see Appendix B.

2. Community Health Status Indicators

The Center for Disease Control (CDC) designed the Community Health Status Indicators (CHSI) study to promote healthy lifestyles in local communities. The CHSI provides online access to important issues affecting a local community's current and future health status (CHSI, 2016). This dataset contains 579 factors for every county in the United States. Factors range from demographic, economic, disease and death, and additional predictors of future health such as obesity and access to healthy foods (Centers for Disease Control and Prevention, 2010). We select 18 factors for inclusion in this research, including suicide rates, obesity rates, and recent drug use based on their resemblance to factors found to be significant in other studies. A full listing of the 18 variables we use is in Appendix B.

3. Integrated Postsecondary Education Data System

The Integrated Postsecondary Education Data System (IPEDS) contains 70 variables for 7,688 postsecondary education institutes. Institutions represented in this data set include community colleges, traditional four-year degree granting universities, seminaries, and trade schools. The data set depicts school size, degrees offered, location, admissions, and educational scores (National Center for Education Statistics, 2016).

Intrater found a negative correlation between the number of universities in or near a ZIP code and the recruit production of that ZIP code (Intrater, 2015).

4. Crime

Crime data comes from the U.S. Department of Housing and Urban Development (HUD) who derived the data from the Federal Bureau of Investigation's (FBI) uniform crime reports (Housing and Urban Development, 2016b). The data contains counts of violent and non-violent crime incidents at the CBSA level for the year 2010. For a more detailed explanation of the importance of crime data in recruiting production models, see Intrater (2015).

B. METHODOLOGY

We use Poisson generalized linear models (GLMs) and gradient boosted decision trees to gain insight from the data. We use the R statistical software program for descriptive statistic calculations and model developments in this chapter (R Core Team, 2016).

1. Poisson GLM

We choose Poisson GLMs for use in this research because of their applicability for count data. A Poisson GLM has three components: a linear predictor similar to a standard linear model, a link function that describes how the mean relates to the linear predictor, and a variance function that describes how the variance relates to the mean (Faraway, 2006).

We use the logarithm of market segment population size as an offset for the model so that we can model accessions as a rate. An offset normally makes use of an exposure variable, which indicates the number of times an event could happen over the exposure period; for the purposes of this research we are using CBSA population size as the exposure variable. By treating population size as an offset we model the CBSA accession rate while maintaining a count response variable for the model (Faraway, 2006).

To provide a model with the fewest number of variables and the most explanatory power, we use a regularized Poisson regression using the lasso or L1 norm of the dependent variable coefficients and implemented using the *cv.glmnet* function from the *glmnet* package in R (Friedman, Hastie & Tibshirani, 2010). *Cv.glmnet* employs algorithms that use cyclical coordinate descent fitting the entire lasso regularization path and then cross-validating the model over k-folds; we use $k = 10$ (Friedman, Hastie, & Tibshirani, 2010). Once the initial model is fit, we check the structural fit of the model by fitting a general additive model where the linear partial fit for each numeric predictor is replaced by a smooth nonparametric fit. We then plot the partial residual plots for each numeric predictor variable. A visual inspection of the plots tells us if the model structure is sound or not. Next we check the variance structure and check for over dispersion. We plot the estimated variance against the mean and estimate our over dispersion parameter (Faraway, 2006). The estimated variance should be proportional to the mean; an increase in the mean should cause an increase in the variance (Faraway, 2006).

To determine the proportion of deviance between CBSAs explained by the model we use a pseudo R-squared value. We compute the pseudo R-squared value as one minus the ratio of the model residual deviance to the null deviance (Faraway, 2006).

2. Gradient Boosted Decision Trees

A gradient boosted decision tree is a regression tree or a classification tree that is grown incrementally to improve prediction results of a tree model. Decision trees are prone to grow large and over-fit the data. Boosting leverages many small trees that are grown sequentially, each tree learning from the previous tree. The small trees learn by fitting to the previous residuals instead of to a response variable. The updated tree is added into a fitted function, which then updates the residuals, and the process continues until a specified number of trees are created (James, Witten, Hastie, & Tibshirani, 2015).

There are three tuning parameters for the gradient boosted decision tree models: the number of trees to grow, the shrinkage or learning rate (lambda), and the number of splits for each tree. A large number of trees can lead to over-fitting of the model. To counter this effect, we use cross validation to select these tuning parameters. The number of splits, *d*, in a model describes the model's level of complexity. If *d* = 1 then the model is additive and the trees are known as stumps. For *d* greater than or equal to 2, the models represent two-way and greater variable interactions (James et al. 2015).

We use the *gbm* function from the gbm package in R to develop our gradient boosted decision trees and we use the *train* function from the *caret* package in R to find the optimal tuning parameters for the *gbm* object (Kuhn, Wing, Weston, Williams, Keefer, Engelhardt, Cooper, Mayer, & Kenkel, 2016). First, we enter a range of tuning parameters for the *train* function to find the best model parameters to maximize the area under the receiver operating characteristic (ROC) curve. The *train* function returns an object that contains the performance values for each combination of model parameters specified. To choose the optimal model, we use ten-fold cross-validation to find the model with the fewest variables that are within one standard error of the best model. This provides us with a model that classifies well with the lowest number of variables. Once we fit the model using the optimal tuning parameters we inspect the model to determine the top influential factors and then test the performance using the *predict* function to develop confusion matrices and ROC curves to evaluate model performance.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. ANALYSIS

Chapter IV contains the models developed to predict USAREC region accessions counts by PRIZM NE market segment. The models we develop provide insights about which factors affect regional market segment production. The first section contains an analysis of the national top producing PRIZM NE market segments. Then, later sections provide deeper analysis of socio-economic factors that affect PRIZM NE market production in CBSAs.

A. IMPACT OF REGION ON THE NUMBER OF ACCESSIONS

In Section A, we fit Poisson GLMs to determine if region affects the number of accessions. For this task, the response variable is the number of accessions for each CBSA and PRIZM NE market segment. We fit two Poisson GLMs: one that contains a categorical variable indicating CBSA and a categorical variable indicating PRIZM NE market segment and one that only contains the PRIZM NE market segment. For both models, the CBSA PRIZM NE market segment population size is used as an offset. We then compare the pseudo R-squared value to determine how much the model deviance is reduced by including CBSA. The following sections describe this process in more detail.

1. Model Development and Variable Selection

We repeat the analysis twice: once for all 66 PRIZM NE market segments and once for only the top 13 accessing PRIZM NE market segments. We create two data frames: one for the top accessing PRIZM NE market segments, and a larger data frame containing all the PRIZM NE market segments. The response variable is the number of accessions; the independent variables are CBSA, PRIZM NE market segment, and PRIZM NE market segment population size for that CBSA. For this section we define an observation as a unique CBSA and PRIZM NE market segment pairing. The number of observations for each data frame is obtained by taking the number of CBSAs times the number of PRIZM NE market segments (13 or 66). During the initial exploration of the models, we observe and then remove a significant number of observations whose population size (for that CBSA, PRIZM NE market segment combination) is zero. Table 4 shows the impact of observation removal from the dataset.

Next, we fit the two Poisson GLM models with both the offset log of PRIZM NE market segment population size and market segment membership, and one that also includes CBSA identity.

Table 4. Removal of Observations with Zero PRIZM NE Market Segment Population.

	Top Market Segments	% of Total	All Market Segments	% of Total
Original Number of Observations	11869	100	60258	100
Number of Zero Population Observations	2948	25	25901	43
Number of Observations Used	8921	75	34357	57

2. Region Poisson GLM Analysis

Table 5 demonstrates that the pseudo R-squared value of the Poisson GLM with CBSA as a predictor was .294 greater than the pseudo R-squared value for the Poisson GLM that did not include CBSA as a predictor. The difference in value means that the Poisson GLM with the CBSA predictor explains 29 percent more of the deviance than a model without a CBSA predictor (Faraway, 2006). We conclude that CBSAs make a contribution to the prediction of the number of Army accessions. We confirm the findings with a large sample likelihood ratio test using the *drop1* function from the STATS package in R, that the categorical variable indicating CBSA cannot be removed from the model, the p-value is less than 2.2e-16 (R Core Team, 2016). While this model's purpose is not to predict accessions at the CBSA level, the model does provide a foundation for additional analysis. The next section explores which CBSA socio-economic factors influence PRIZM NE market segment accession.

Table 5. Pseudo R-squared Results for Top Market Segments and All Market Segment GLM Models.

	Top 13 Market Segments	All Market Segments
Without CBSA	0.1603	0.2117
With CBSA	0.4543	0.4182
Increase in Pseudo R-squared	**0.294**	**0.2063**

B. IN-DEPTH ANALYSIS OF PRIZM NE SEGMENT 32 ACCESSIONS

1. Variability of PRIZM NE Segment 32 Accessions Rate

To explore the regional factors that might affect PRIZM NE market segment accessions, we isolate a high accessing PRIZM NE market segment with a high market penetration rate, Segment 32. First we examine Segment 32 to determine if the segment has the same accession rate across regions. Next we will use Segment 32 to explore the socio-economic factors that contribute to accession rate variability across CBSAs. We map the number of accessions per 100,000 people with ArcGIS and conduct a visual analysis of Segment 32's accession rate. Figure 3 shows the differences in Segment 32 accessions by CBSA and general regions of the country. Accession rates are lowest in the Northeast and the Northern Midwest, while the Southern Regions show higher rates of accessions. When we compare Segment 32 to Segment 20, another high accessing PRIZM NE market segment, their accession rate patterns obviously differ. The Northeast region of Nevada, the Northeast region of Wyoming, and South Dakota all demonstrate three easily identifiable examples of the different accession rates between PRIZM NE segments. Figure 4 shows the number of accessions per 100,000 people for PRIZM NE Segment 20.

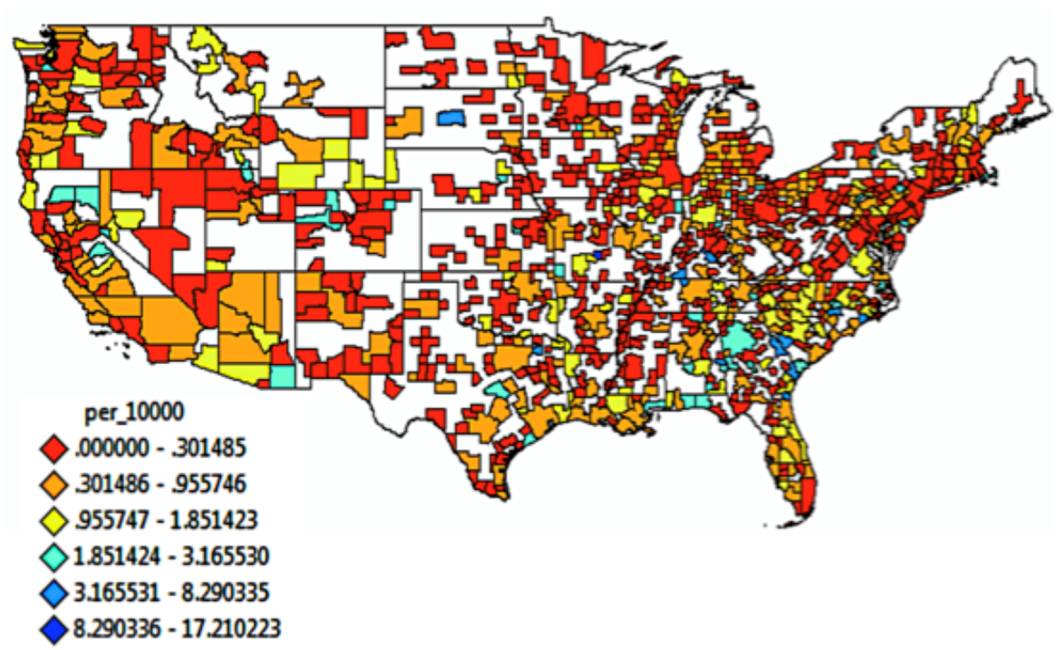


Figure 3. PRIZM NE Segment 32 Accessions per 100,000 People from 2013 USAREC PRIZM NE Data.

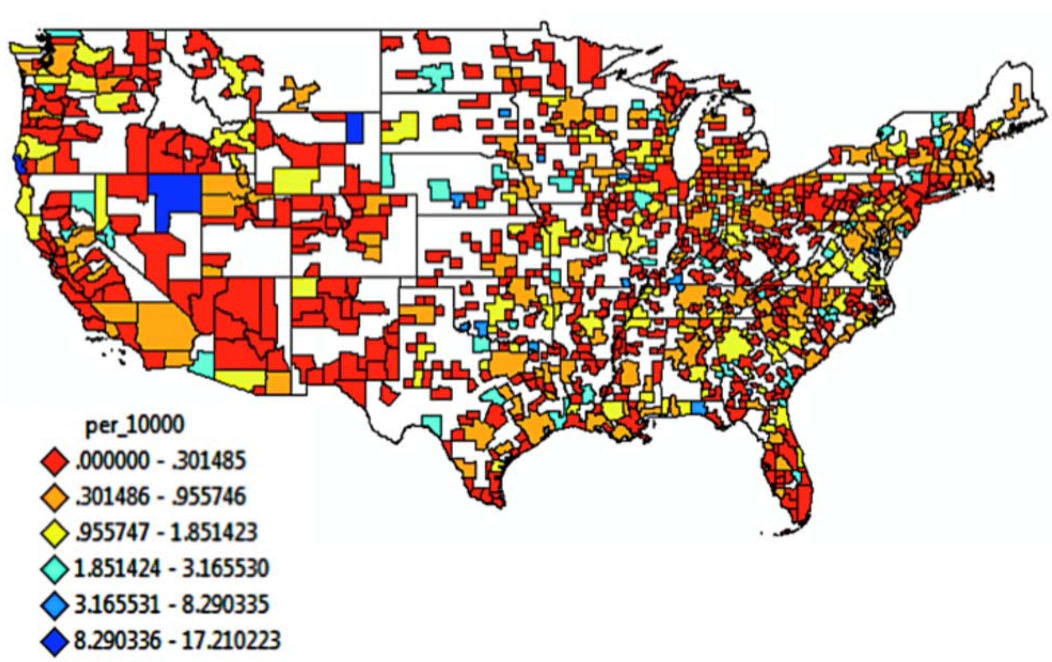


Figure 4. PRIZM NE Segment 20 Accessions per 100,000 People from 2013 USAREC PRIZM NE Data.

We test the null hypothesis that PRIZM NE Segment 32 accession rates are constant for all CBSAs using the Chi-squared test for homogeneity; we find evidence to reject the null hypothesis that Segment 32 accession rates are constant across CBSAs at the five percent significance level (p-value < 2.2e-16). Therefore, we conclude that PRIZM NE Segment 32 rates do vary by region. To find whether this remains true even when CBSAs have comparable population size, we conduct the Chi-squared test on samples of CBSAs with similar population sizes. We find that we can reject the null hypothesis that Segment 32 accession rates are the same for CBSAs with population sizes of less than 500,000 people (p-value < 2.2e-16) but cannot reject the null hypothesis for CBSAs with population sizes greater than or equal to 500,000 (p-value 0.93). Table 6 shows a sampling of CBSAs that have similar population sizes and have statistically different Segment 32 accession rates. We see that the Killeen-Temple, TX CBSA may be an outlier: its accessions are almost two times greater than the CBSA with the next-largest number of accessions. The presence of the largest Army base in the United States may be a confounding factor.

Table 6. The Number of Accessions for CBSAs with Population Sizes Between 400,000 and 500,000 that also have Sizeable Segment Population Sizes.

CBSA	CBSA Name	Accessions per 100,000	Accessions	Segment Population	CBSA Population
28660	Killeen-Temple, TX	5.257	24	1690	456533
23580	Gainesville, GA	2.858	13	1934	454885
15940	Canton-Massillon, OH	1.210	6	1493	495910
33660	Mobile, AL	1.142	5	1746	437726
25180	Hagerstown-Martinsburg, MD-WV	0.968	4	964	413077
12100	Atlantic City–Hammonton, NJ	0.862	4	1629	463868
37900	Peoria, IL	0.493	2	946	405600
18580	Corpus Christi, TX	0.433	2	1323	461871

2. Poisson GLM for Regional Factors That Affect Segment 32 Accessions

The analysis of the Poisson GLM takes into account the model's goodness of fit, model assumptions, and what the model's structure implies about a region's effect on PRIZM NE Segment 32 recruit production.

The pseudo R-squared value of the final Poisson GLM was 0.52. This value indicates that the Poisson GLM can explain 52 percent of the deviance of the CBSA Segment 32 accession rates model fit (Faraway, 2006).

There is no evidence from partial residual plots of the numeric variables to indicate that the model structure is not sound or that any of the numeric variables need to be transformed. To view the partial residual plots please see Appendix D. The plot of estimated variance vs. mean plot, Figure 12 in Appendix D, looks unusual because of the high number of zero accessions, but the variance appears to increase as the mean value increases as we expect.

We then check for over-dispersion. The normal dispersion parameter (ϕ) for Poisson regression is ϕ = 1; values of ϕ greater than one mean over dispersion and for values less than one mean under dispersion (Faraway, 2006). We estimate the dispersion parameter for the Segment 32 Poisson to be 1.78 indicating that our model is slightly over-dispersed. To account for over dispersion, we fit an over-dispersed Poisson GLM which adjusts standard errors and tests of hypothesis to accommodate the extra variability in the response variable (Faraway, 2006).

The structural composition of the Poisson GLM provides insight into how regional factors might influence a CBSAs PRIZM NE Segment 32 recruit production. We use Poisson regularized regression to remove variables. To analyze the structural composition, the first step is to inspect which variables the model retains along with the coefficient values. Table 7 contains the predictor variables, coefficients, and standard error of the model. The following paragraphs discuss the analysis of the variables retained in the over dispersed Poisson GLM. A negatively signed coefficient relates to a decrease in Segment 32 accession rates and a positively signed coefficient relates to an increase in Segment 32 accession rates if all other variables could be held constant.

Table 7. PRIZM NE Segment 32 Poisson GLM Coefficients and Standard Errors (SE).

Predictor Variable	Coefficient	SE
The death rate per 100,000 people	-4.47E-01	3.73E-03
The % of adults who have diabetes	7.17E-02	3.26E-02
The % of adults who smoke	6.26E-02	1.18E-02
The number of post-secondary education schools	4.41E-02	9.19E-03
The % of adults at risk to health issues related to lack of exercise	-3.98E-02	1.17E-02
The % of adults who receive less than recommended amounts of fruit and vegetables	-2.23E-02	1.17E-02
The rate of stroke related deaths per 100,000	1.09E-02	5.25E-02
The CBSA median income	-4.35E-05	6.08E-06
The number of people that work in an arts related job	-2.16E-05	3.93E-06
The number of people with no family or roommates	-1.30E-05	2.11E-06
Youth with some college (18 to 24 year olds)	-1.01E-05	3.71E-06
The number of veterans	8.50E-06	1.78E-06
The number of armed forces members employed	8.40E-06	4.65E-06
The number of people unemployed	7.87E-06	1.62E-06
The number of non-violent crime incidents	-5.58E-06	2.06E-06
The number of vacant housing units	5.40E-06	1.54E-06
The number of people who work in public administration job	-5.33E-06	1.92E-06
Youths that did not graduate high school (18 to 24 year olds)	1.58E-06	9.17E-06
The number of people that commute30 to 60 minutes	2.09E-07	3.32E-08

There appear to be two categories of predictor variables that negatively influence PRIZM NE Segment 32 accession rates: variables that indicate regions of increased economic opportunity and variables that indicate a decrease in the available pool of recruit candidates. We note that these coefficients represent the partial effect of each variable if all other variables could be held constant. An increase in the number of 18 to 24 year-olds that have some college, the number of people employed in the arts, the number of people employed in public administration, and the median income all represent economic incentives for potential recruits to join the military. Variables such as increase in the total number of deaths, the number of people who do not exercise, and the number of people that have poor diets decrease the available recruiting pool. The direction of the effects of these variables aligns with expectation. Contrary to expectation, an increase in non-violent crime relates to a decrease in the accession rate. Intrater (2015) found that an increase in non-violent crime within a ZIP code was a good indicator of an increase in overall Navy recruit production for that area.

Among the predictor variables that might positively influence the PRIZM NE Segment 32 accession rate for CBSAs, are those that may indicate a higher proportion of lower income people. Also, a factor that may indicate how important military presence is for recruiting. Increases in the number of people who do not have a high school diploma, the number of people who are unemployed, the percentage of people who smoke, the percentage of people with diabetes, and the number of vacant homes in an area seem to increase the accessions rate for a CBSA. An increase in these variables may indicate that a region has limited economic opportunity (Metcalf, Scragg, Schaaf, Dyall, Black, & Jackson, 2008). An increase in the presence of veterans and military personnel that are employed in an area also increase the predicted CBSAs accession rate. A high veteran population in a given area means that veterans may have more opportunities to demonstrate the positives of a military career, and veteran participation could mean a positive influence in accession rates. The number of armed forces personnel employed in an area could represent the presence of a military base or the number of recruiters in a region, both of which may have a positive influence on accession rates.

The over dispersed Poisson GLM provides many insights into how a CBSA's regional factors influence PRIZM NE Segment 32's accession rate. While the model is likely not robust enough to predict recruiting center accessions, it does provide a basis for further analysis.

3. Gradient Boosted Decision Tree Model for Regional Factors that Affect Segment 32 Accessions

The analysis of the gradient boosted decision tree model focuses on the model’s goodness of fit, and what the model’s structure implies about a region’s effect on PRIZM NE Segment 32 recruit production.

We construct two new variables. We first construct a PRIZM NE Segment 32 accession rate variable, the number of accessions per 100,000 people. We then construct a binary response variable one for CBSAs that have accession rates at or above the mean accession rate and zero for CBSAs that have accession rates below the mean. The binary response variable allows us to fit models to classify CBSAs as either performing better or worse than the average accession rate. We test our model performance on training and test sets; the results are listed in subsequent paragraphs.

The results in Table 8 show that the boosted tree model fit produces a correct classification rate of 75 percent for the training set and a correct classification rate of 70 percent for the test set. The gradient boosted decision tree model had a higher classification rate of 95 percent for the training set and 91 percent for the test set on CBSAs that have accession rates below the mean accessions rate. Conversely, the model has a lower classification rate of 38 percent for the training set and 30 percent for the test set for those CBSAs that have accession rates above the mean accession rate.

Table 8. Confusion Matrix for Gradient Boosted Decision Tree Model.

Training Set Confusion Matrix

		Predicted Value		
		0	**1**	**Correct Classification Rate**
Actual Value	**0**	385	18	0.955
	1	140	88	0.386
Overall Correct Classification Rate				**0.750**

Test Set Confusion Matrix

		Predicted Value		
		0	**1**	**Correct Classification Rate**
Actual Value	**0**	384	35	0.916
	1	153	65	0.298
Overall Correct Classification Rate				**0.705**

The receiver operating characteristics (ROC) curve plot offers additional information on the gradient boosted decision tree (James et al. 2015). The ROC plot depicted in Figure 5 illustrates the tradeoff between the true positive rate and the false positive rate. An analyst can choose the observation classification cutoff point, which will vary the true positive and false positive rates. As a reference, the confusion matrices for the gradient boosted decision tree displayed in Table 8 uses a cutoff value of 0.5. The area under the curve (AUC) of the ROC plot gives the overall performance of a classifier over all possible outcomes. An AUC of one denotes that the classifier does a perfect job of classifying observations as opposed to an AUC of 0.5 which means the classifier is comparable to flipping a coin when classifying an observation (James et al. 2015). The gradient boosted decision tree model produced an AUC of 0.83 for the training set and 0.72 for the test set.

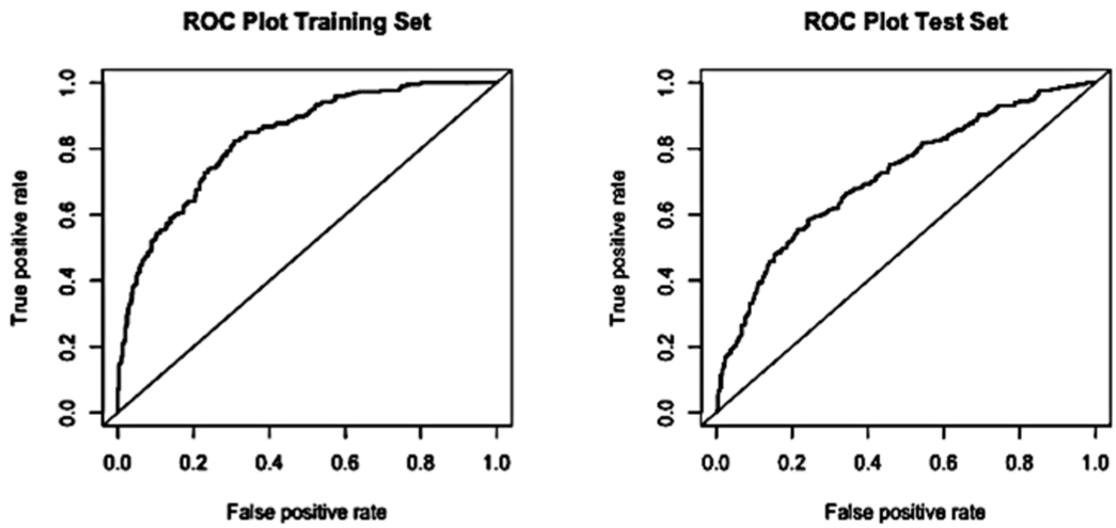


Figure 5. PRIZM NE Segment 32 Gradient Boosted Decision Tree Model ROC Curves for Training and Test Sets.

The accuracy (where accuracy is the proportion of observations correctly classified) vs. cutoff plot depicted in Figure 6 provides information on how changing the cutoff point will impact the accuracy of the model predictions (Zhoa & Cen, 2014). Figure 6 shows that there is an area between cutoff values of 0.3 and 0.6 were the model experiences moderate changes in performance. This indicates that an analyst could alter the cutoff point between these ranges and experience little drop off in accuracy. This shows that gradient boosted decision tree model can provide the user with some flexibility in using the model without sacrificing accuracy.

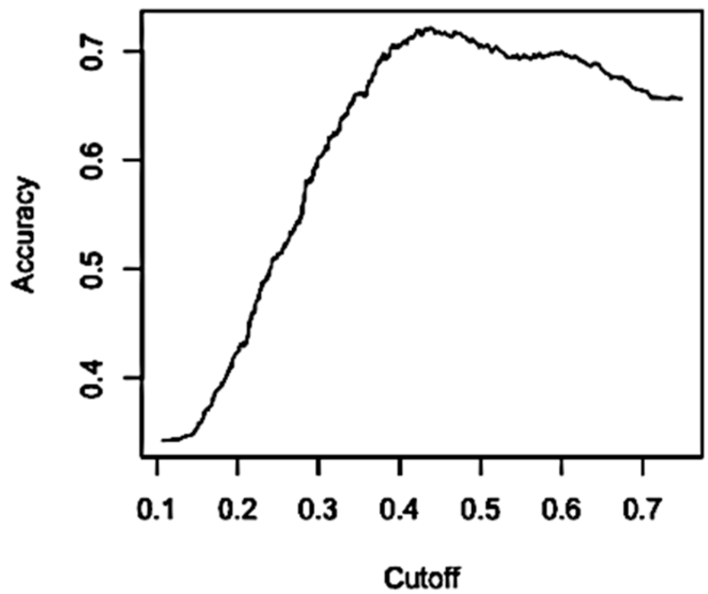


Figure 6. PRIZM NE Segment 32 Gradient Boosted Decision Tree Model Accuracy vs. Cutoff Plot.

The importance that the gradient boosted decision tree model places on each of the predictor variables provides significant insight. The *gbm* function also returns the relative influence of each variable in reducing the loss function used to fit the model (Ridgeway, 2015). The *gbm* function scales the influence values so that they sum to 100. The variable influence metric provides insight into the influence each predictor variable has on the model. Table 9 shows the top 10 variable influence values for our model. The gradient boosted decision tree finds several regional factors important that were also important to the Poisson GLM. These variables include the number of armed forces personnel employed in a CBSA, the total number of deaths in a CBSA, and the median CBSA income. Similar to the Poisson GLM, the gradient boosted decision tree model finds that an increase in the percentage of people that smoke and have diabetes may lead to an increase in the CBSA's accession rate.

Table 9. Top 10 Influential Variables for the PRIZM NE Segment 32 Gradient Boosted Decision Tree Model.

Predictor Variable	Influence	Direction
The number of armed forces members employed	14.24	Positive
The death rate per 100,000 people	11.75	Negative
The CBSA's median income	6.99	Negative
The % of adults with high blood pressure	5.55	Positive
The per capita income	5.31	Negative
The % of adults with diabetes	4.68	Positive
The number of people that work in an agriculture related job	4.42	Negative
The % of adults who smoke	4.40	Positive
The number of deaths by suicide per 100,000 people	4.15	Negative
The number of Major Depressive episodes per 100,000 people	3.87	Negative

C. ANALYSIS OF TOP 5 PRIZM NE SEGMENTS AND SEGMENT 47 ACCESSIONS

1. Poisson GLM Regional Factors that Affect the Top Five PRIZM NE Segments

We now fit Poisson GLM for the remaining five highest producing PRIZM NE market segments and Segment 47 from the 2012 data set to find common predictor variables of top accessing PRIZM NE market segments. We also fit a Poisson GLM for a market segment that USAREC has requested. The PRIZM NE market segments that we model are Segment 20 (Fast Track Families), Segment 33 (Big Sky Families), Segment 37 (Mayberry-ville), Segment 47 (City Startups), and Segment 63 (Family Thrifts). In addition to building models for the additional market segments, we apply the Poisson GLM built for PRIZM NE Segment 32 to test its applicability across market segments. Common predictor variables in most of the models include the number of armed forces members employed in a CBSA, the number of veterans in a CBSA, the volume of non-violent crime, and the total number of deaths in a CBSA. For a complete comparison of the models see Appendix E. Table 10 displays the model fits for the top accessing PRIZM NE market segments and PRIZM NE Segment 47. In most cases the Poisson GLM fitted for the PRIZM NE Segment 32 performs better than the models fit specifically for each market segment. For five of the six models we are able to explain over 30 percent of the deviance of PRIZM NE market segment accession rates.

Table 10. 2012 PRIZM NE Top Segments Poisson GLM Performance.

PRIZM NE Segment	Pseudo R2 for Segment Model	Pseudo R2 Segment 32 model
Segment 32	0.521	NA
Segment 20	0.219	0.262
Segment 33	0.380	0.379
Segment 37	0.313	0.326
Segment 47	0.358	0.372
Segment 63	0.351	0.426

2. Gradient Boosted Decision Tree Model of Regional Factors that Affect the Top Five PRIZM NE Market Segments

We fit the gradient boosted decision tree models for each of the remaining top five PRIZM NE market segments and Segment 47. Comparing the model results from the Poisson GLM to the gradient boosted decision tree we note that for each market segment at least two variables that appear in the market segment's Poisson GLM also appear in the market segment's decision tree. The gradient boosted decision tree did not identify several factors that were important in Poisson GLMs, such as the Veteran population of a CBSA and the volume of non-violent crime. The top five factors found most prevalently in models include armed forces employment, total deaths, the number of agricultural workers, per capita income, and the number of people who die from lung cancer. For a detailed list of each market segment's most influential decision variables see Appendix E, Table 16. Gradient boosted decision trees AUC across PRIZM NE market segments is reasonably consistent as seen in Table 11. A possible way to operationalize these models would be for USAREC to create recruiter reference sheets to augment the JAMRS marketing guide that highlight the regional characteristics that contribute to or detract from the recruiting centers' accession rates.

Table 11. Top 5 PRIZM NE Market Segments and Segment 47 Gradient Boosted Decision Tree AUC.

	AUC Train	**AUC Test**
Segment 32	0.833	0.718
Segment 20	0.874	0.646
Segment 33	0.832	0.690
Segment 37	0.847	0.673
Segment 47	0.884	0.689
Segment 63	0.792	0.701

THIS PAGE INTENTIONALLY LEFT BLANK

V. CONCLUSION AND RECOMMENDATIONS

Chapter V provides a summary of the analytic approach and results discussed in the previous chapters along with recommendations for USAREC and areas for future research.

A. SUMMARY

The goal of this thesis is the development of statistical models that could aid USAREC in identifying if and why PRIZM NE market segment recruit production varies by region. We fit models to assess regional factors that may affect PRIZM NE market segment accession rates. We pose two questions for analysis to help to accomplish this goal:

- Do PRIZM NE market segments produce the same number of recruits per capita regardless of region?
- Within a PRIZM NE segment, which regional factors affect recruit production?

The expectation before this research was that PRIZM NE market segments would produce the number of accessions at a homogeneous rate regardless of region, because of the detailed categorization that PRIZM NE represents. We find instead that PRIZM NE market segments are not homogeneous in accessions rates.

We conclude that the regional factors that affect recruit production vary among the top five PRIZM NE market segments, but some factors are prevalent among the market segments. In general terms, the models find that an increased presence of military personnel in a region, earning potential, and public health affects regional accession rates. Specifically, the number of members of the armed forces in a CBSA positively impact accessions rates, while the median income of the CBSA, and the total death rate per 100,000 people negatively impact accession rates. These are the top three regional factors that affected a region's accession rates.

To support the development of statistical models, this research aggregates open source data from six different agencies into two data sets with a combined total of 139 variables for 910 CBSAs. We develop two models, including a Poisson GLM and a gradient boosted decision tree, to provide USAREC G2 analysts with the information necessary to better access recruiting markets.

The research finds that both Poisson GLMs and gradient boosted decision trees can be used to explore the regional factors that affect PRIZM NE market segment recruit production and also that gradient boosted decision trees can predict a CBSAs ability to under or over-perform compared to the mean PRIZM NE market segment accession rate. These models provide USAREC with a methodology to access market segment accession rate variability, which they previously lacked.

Future research regarding recruitment should include focusing on county boundaries instead of CBSAs and also a cluster analysis of county-level socioeconomic factors comparing county clusters to PRIZM NE market segmentation. By addressing these two areas, it is possible that USAREC G2 analysts will achieve greater recruiting market understanding by augmenting existing model structures.

B. RECOMMENDATIONS

We recommend that USAREC implement gradient boosted models as decision support tools for identifying recruiting market potential in an objective and repeatable manner. The employment of gradient boosted decision trees allows USAREC to identify areas, scalable to any level of geography, with high probabilities of supporting the recruiting mission. Additionally, the model could be modified to identify “must keep,” “must win,” and “markets of opportunity” ZIP codes. These models could be used independently or in concert with the current SAMA model.

APPENDIX A. DATA CLEANING

Appendix A provides the reader with information used in the development and transformation of the raw data into the final format used in the model.

To aggregate ZCTA level data up to CBSA level data, we use shape files and a ZCTA to CBSA relation file from the U.S. Census Bureau’s website United States Census Bureau (2016b). ArcGIS is a software package that allows us to associate data to physical spaces and combine, separate, or apportion it as necessary. If a ZCTA intersects with more than one CBSA, the data is distributed to the CBSAs based on the proportion of space of the ZCTA within each CBSA.

To convert county level data to CBSA level data we first translate the data from county to ZIP code level then from ZIP code to CBSA level. We use crosswalk files from the HUD website for these translations (Housing and Urban Development, 2016a). We use the 2015 crosswalks because we are interested in geographic boundaries as they are today, as opposed to when HUD originally collected the data. The crosswalks use the portion of residential addresses in each ZIP code within the county to apportion the data from county to ZIP code. We use the same process to translate from ZIP code to CBSA level.

Table 12. RISDs Excluded From the PRIZM NE Data Files Before Being Aggregated to CBSA Level.

RSID	Region
1A8G	Europe APO
1A8H	Europe APO
1A8J	Europe APO
1A8M	Europe APO
3G6C	Puerto Rico
3G6D	Puerto Rico
3G6E	Puerto Rico
3G6G	Puerto Rico
3G6H	Puerto Rico
3G6J	Puerto Rico
3G6M	Puerto Rico
3G7A	Puerto Rico
3G7G	Puerto Rico
3G7M	Puerto Rico
3G7P	Puerto Rico
3G7R	Puerto Rico
3G7S	Puerto Rico
3G7T	Puerto Rico
3G7V	Puerto Rico
6H7B	Armed Forces Pacific
6H7G	Armed Forces Pacific
6H7J	Armed Forces Pacific
6H7K	Armed Forces Pacific
6H7N	Armed Forces Pacific

APPENDIX B. META DATA

Appendix B provides meta-data that explains each variable.

Table 13. Meta-Data for CBSA-level Data.

Variable	Description
CBSA	Core Based Statistical Area code
CBSA_Name	CBSA name
Population	CBSA population
Undergrad_Enrollment	Total # undergraduate enrollment in CBSA
Gradstudent_Enrollment	Total # graduate student enrollment in CBSA
Population_18to24	CBSA population of 18 to 24 year olds
enrolled_college_18to24	Total # of college students 18–24 enrolled in CBSA
non_hs_grad_18to24	Total # 18–24 year olds that did not graduate high school
hs_grad_18to24	Total # of 18–24 year olds highest level of education is high school diploma
somecollege_18to24	Total # of 18–24 year olds with some college experience
college_grad_18to24	Total # of 18–24 year olds with a college degree
Pop_Over25	CBSA population 25 years and older
Over25_hs_grad	Total # of 25 year olds plus, highest level of education is a high school diploma
Over25_college_grad	Total # of 25 year olds plus that have a college degree
employed	Total # of people employed
unemployed	Total # of people unemployed
armed_forces_employed	Total # of people employed by armed forces
agriculture	# of people employed in agriculture, forestry, fishing, hunting, and mining
construction	# of people employed in construction
manufacture	# of people employed in manufacturing
wholesale	# of people employed in wholesale trade
retail	# of people employed in retail trade
transportation	# of people employed in transportation, warehousing, and utilities
information	# of people employed in information services
financial	# of people employed in finance, insurance, and real estate
professional	# of people employed in professional, scientific, management , administrative, and waste management

	services
education	# of people employed in education, health care, and social assistance
arts	# of people employed in arts, entertainment, recreation, food services, and accommodations
public_admin	# of people employed in public administration
HH_income_lessthan25000	# of households that make less than $25,000 in the last 12 months
HH_income_25000to49999	# of households that make more than or equal to $25,000 but less than $50,000 in the last 12 months
HH_income_50000to74999	# of households that make more than or equal to $50,000 but less than $75,000 in the last 12 months
HH_income_75000to99999	# of households that make more than or equal to $75,000 but less than $100,000 in the last 12 months
HH_income_100to199999	# of households that make more than or equal to $100,000 but less than $200,000 in the last 12 months
HH_income_200000plus	# of households that make more than or equal to $200,000 in the last 12 months
percapita_income	Per capita income of CBSA
below_poverty	# of people below the poverty line
above_poverty	# of people above the poverty line
commute_lessthan_30min	# of people that commute less than 30 minutes for work
commute30to60	# of people that commute between 30 to 60 minutes for work
commute_60plus	# of people that commute more than 60 minutes for work
total_households	# of households in CBSA
married_family	# of two parent families
single_parent_dad	# of single parent families with a father only
single_parent_mom	# of single parent families with a mother only
non_fam_alone	# of people with no family living alone
non_fam_roommate	# of people with no family living with roommate
housing_units	# of housing units
occupied_units	# of occupied housing units
vacant_units	# of vacant housing units
Veterans	# of veterans living in CBSA
Homicide	# of homicides
Lung_Cancer	# per 100,000 people that die from Lung Cancer
Stroke	# per 100,000 people that die from Stroke
Suicide	# per 100,000 people that die from suicide
Total_Births	# of births per 100,000 people

Total_Deaths	# of deaths per 100,000 people
No_Exercise	The percentage of adults at risk to health issues related to lack of exercise
Few_Fruit_Veg	The percentage of adults reporting an average fruit and vegetable consumption of less than 5 times per day.
Obesity	The calculated percentage of adults at risk for health problems related to being overweight based on body mass index (BMI). A BMI of 27.8, for men, and 27.3, for women, or more is considered obese.
High_Blood_Pres	The percentage of adults who responded yes to survey question about high blood pressure.
Smoker	The percentage of adults who responded, "yes" to survey question about smoking.
Diabetes	The percentage of adults who responded yes to diabetes survey question.
Uninsured	The estimated number of uninsured individuals
Unhealthy_Days	The average number of unhealthy days (mental or physical) in the past 30 days.
Major_Depression	Estimate of the number of individuals, age 18 and older, experiencing a major depressive episode during the past year.
Recent_Drug_Use	Estimate of the number of individuals, age 12 and older, using illicit drugs within the past year
Toxic_Chem	Toxic release inventory (TRI) data, amount (in pounds) of total chemicals released.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C. TOP ACCESSING PRIZM NE MARKET SEGMENTS

Table 14 details the top accessing market segments as defined by JAMRS. The table description details the economic status, family characteristics, and some educational characteristics.

Table 14. JAMRS Defined Top Accessing Army Market Segments. Adapted from: Joint Advertising Market Research & Studies, 2014; Nielsen, 2016.

Top Accessing Market Segments for U.S. Army		
Segment	Segment Name	Description
13	Upward Bound	Soccer Moms and Dads in small satellite cities, upper-class families boast dual incomes, college degrees and new split-levels and colonials.
18	Kids & Cul-de-Sacs	Upscale, suburban, married couples with children an enviable lifestyle of large families in recently built subdivisions. This segment is a refuge for college-educated, white-collar professionals with administrative jobs and upper-middle-class incomes.
20	Fast-Track Families	With their upper-middle-class incomes, numerous children and spacious homes, Fast-Track Families are in their prime acquisition years.
32	New Homesteaders	Young, middle-class families seeking to escape suburban sprawl find refuge in New Homesteaders, a collection of small rustic townships filled with new ranches and Cape Cods.
33	Big Sky Families	Scattered in placid towns across the American heartland, Big Sky Families is a segment of young rural families who have turned high school educations and blue-collar jobs into busy, middle-class lifestyles.
34	White Picket Fences	Midpoint on the socioeconomic ladder, residents in White Picket Fences look a lot like the stereotypical American household of a generation ago: young, middle-class, married with children.
36	Blue-Chip Blues	Blue-Chip Blues is known as a comfortable lifestyle for young, sprawling families with well-paying blue-collar jobs.

37	Mayberry-ville	Mayberry-ville harks back to an old-fashioned way of life. In these small towns, middle-class couples and families like to fish and hunt during the day, and stay home and watch TV at night.
41	Sunset City Blues	Scattered throughout the older neighborhoods of small cities, Sunset City Blues is a segment of lower-middle-class singles and couples who have retired or are getting closed to it.
45	Blue Highways	On maps, blue highways are often two-lane roads that wind through remote stretches of the American landscape. Among lifestyles, Blue Highways is the standout for lower middle-class couples and families who live in isolated towns and farmsteads.
50	Kid Country, USA	Widely scattered throughout the nation's heartland, Kid Country, USA is a segment dominated by large families living in small towns.
51	Shotguns & Pickups	The segment known as Shotguns & Pickups came by its moniker honestly: it scores near the top of all lifestyles for owning hunting rifles and pickup trucks.
56	Crossroads Villagers	With a population of middle-aged, blue-collar couples and families, Crossroads Villagers is a classic rural lifestyle. Residents are high school-educated, with lower-middle incomes and modest housing; one-quarter live in mobile homes.

APPENDIX D. MODEL DIAGNOSTICS

Appendix D provides the reader with diagnostic plots for validation of model assumptions, and other information relevant to each model's fit.

Figure 12 shows the plot of estimated variance vs. mean plot. Figures 7 through 11 show the partial residual plots generated for each numeric predictor variable within the Segment 32 Poisson GLM by fitting a generalized additive model where the numeric predictors are replaced by smooth nonparametric functions of those predictors estimated at model fitting. Examining the partial residual plots with the smooth partial fits and standard error bars (dashed lines) determine if any of the numeric predictor variables need to be transformed. There are no discernable patterns in the plots; therefore none of these variables require transformation.

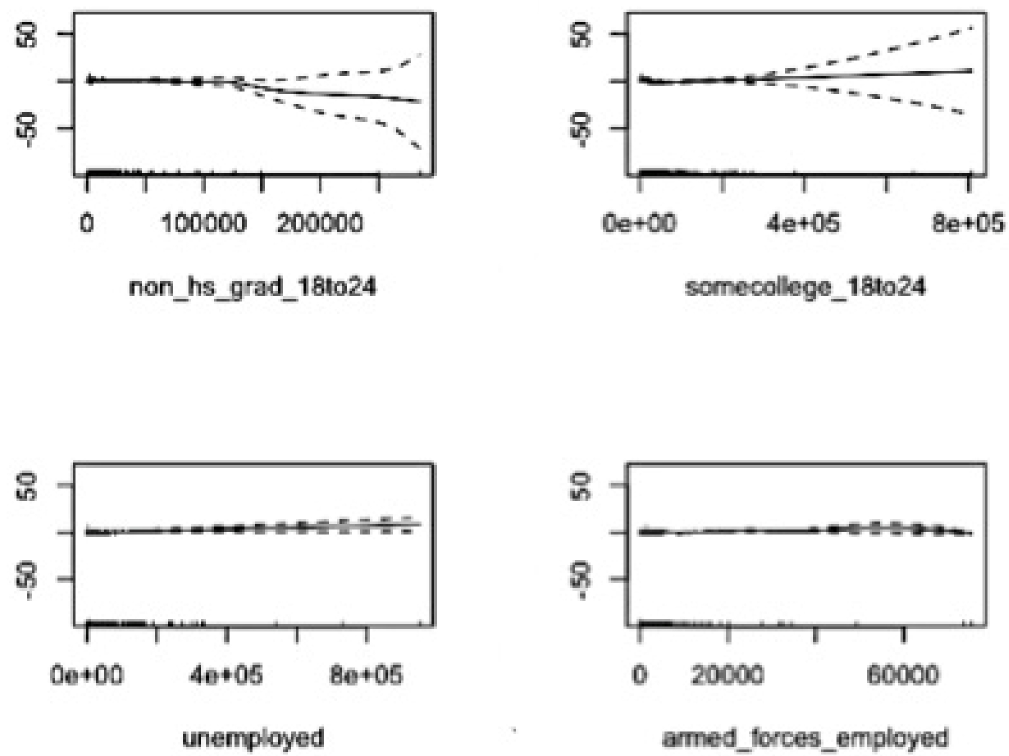


Figure 7. Number of Non-high School Graduates, Number of People with Some College, the Number of People That are Unemployed, and the Number of Armed Forces Members Employed Partial Residual Plot for Segment 32 Poisson GLM.

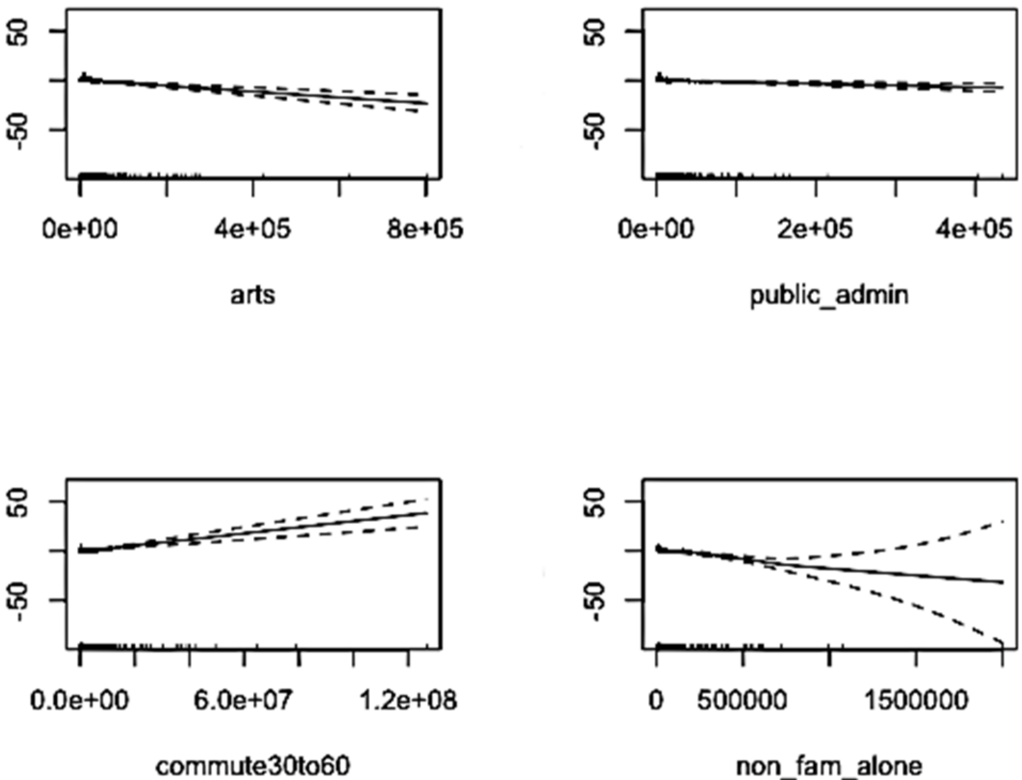


Figure 8. Number of People Employed in the Arts, Number of People Employed in Public Administration, the Number of People who Commute 30 to 60 Minutes, and the Number of People who Live Alone with no Family Partial Residual Plots for Segment 32 Poisson GLM.

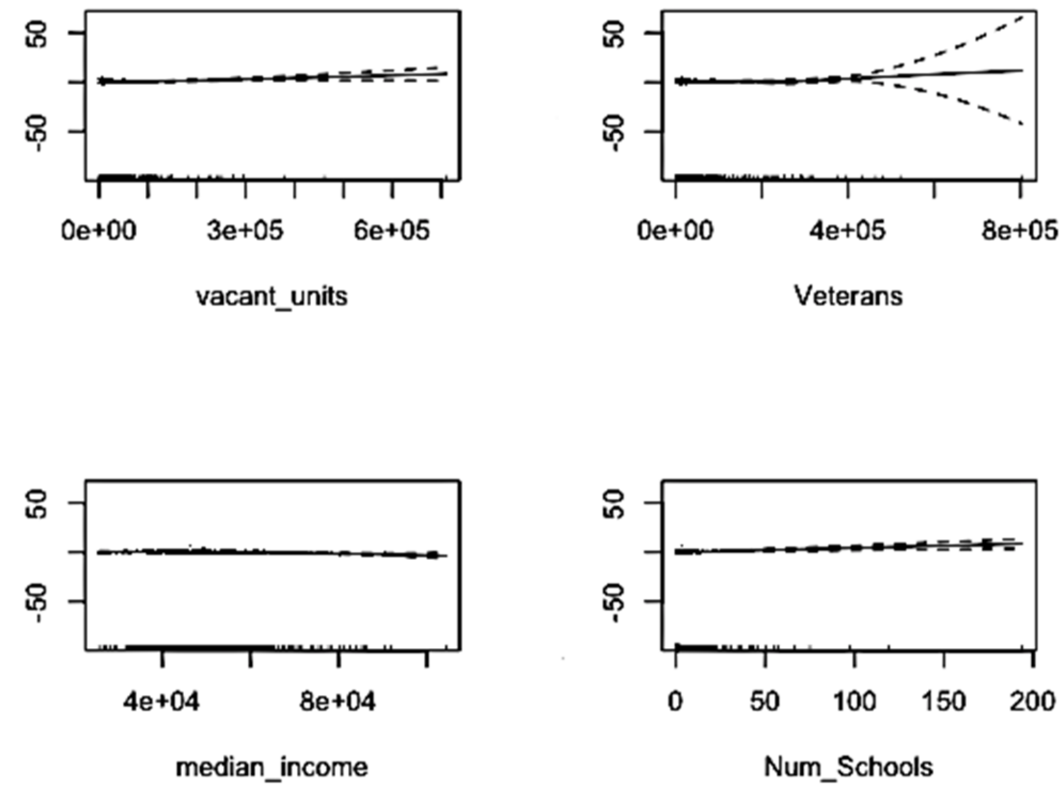


Figure 9. Number of Vacant Units, the Number of Veterans, the Median Income, and the Number of Colleges Partial Residual Plots for Segment 32 Poisson GLM.

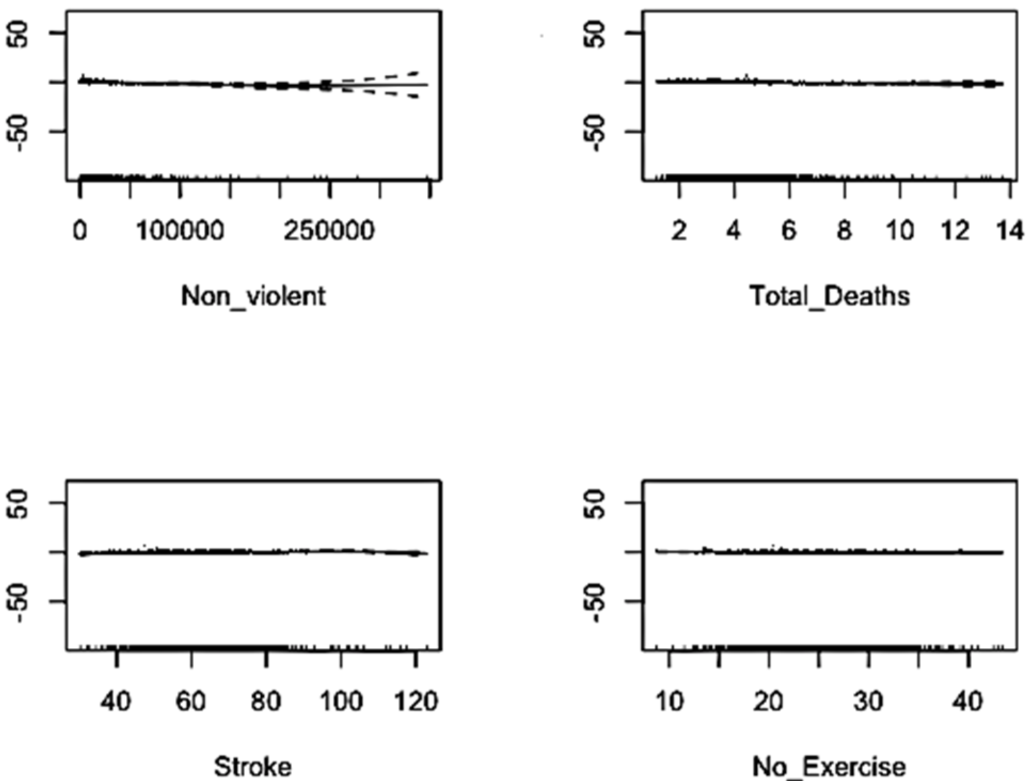


Figure 10. Number of Non-violent Crime, the Total Number of Deaths, the Number of Stroke Related Deaths per 100,000, and the Percentage of Adults at Risk for Health Issues Related to Lack of Exercise Partial Residual Plots for Segment 32 Poisson GLM.

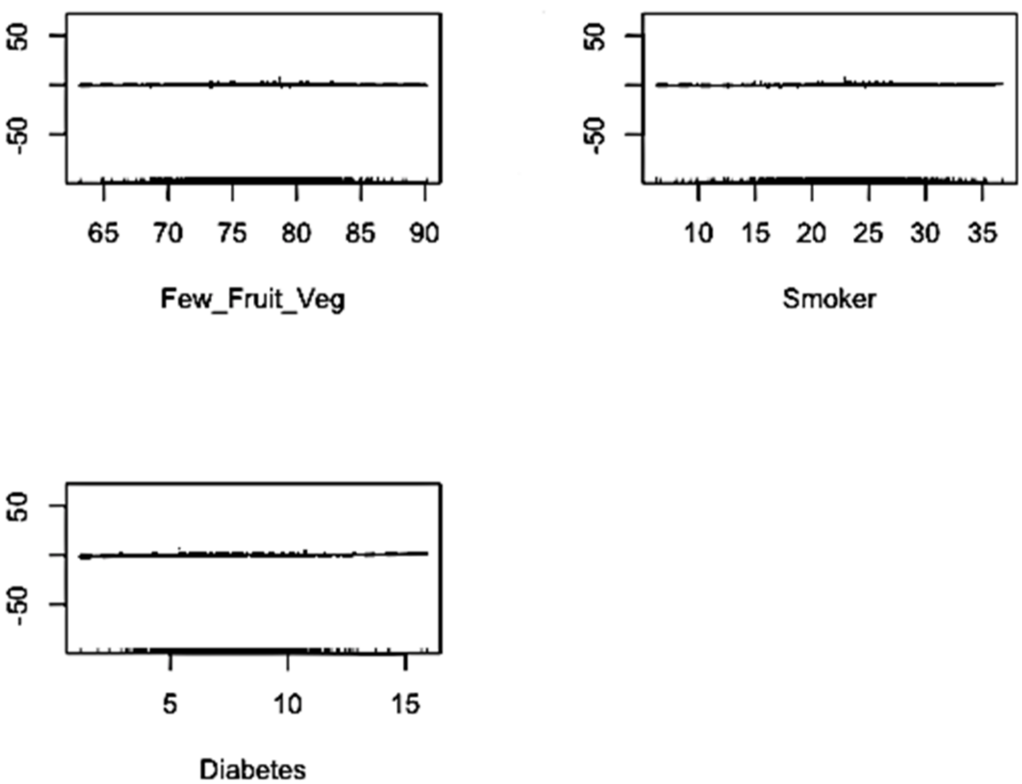
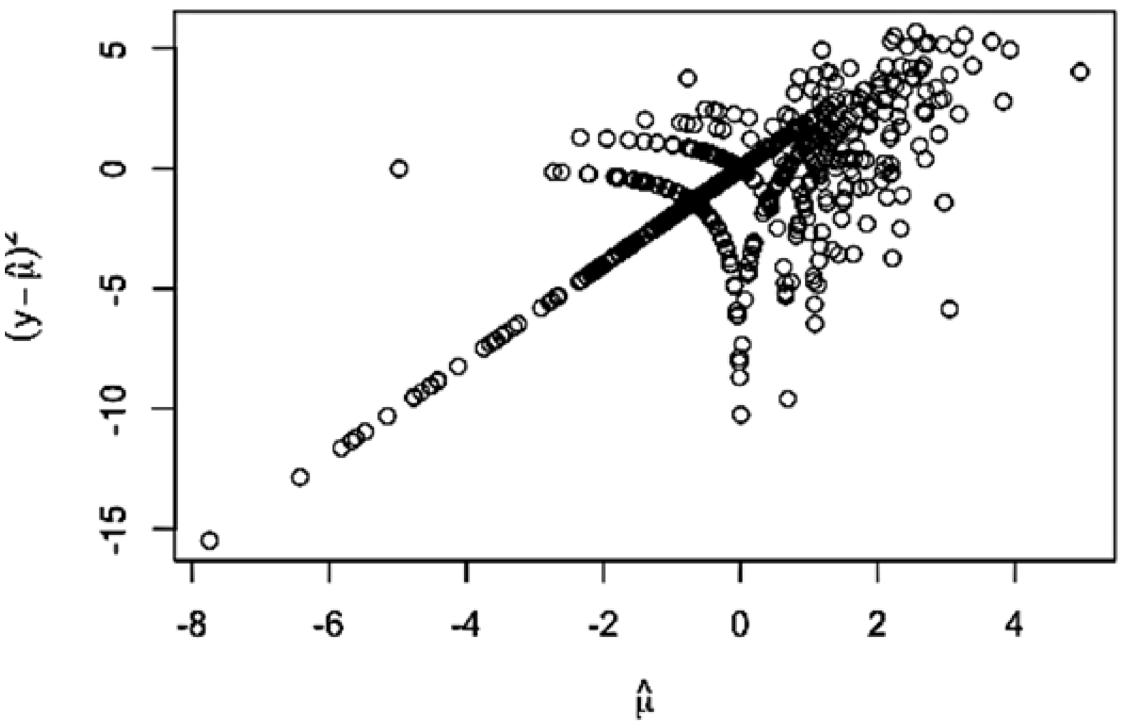


Figure 11. Percentage of Adults That Do Not Receive the Recommended Portions of Fruits and Vegetables, the Percentage of Adults Who Smoke, and the Percentage of Adults with Diabetes Partial Residuals Plots for Segment 32 Poisson GLM.



The plot shows the estimated variance (y-mu hat)^2 against the estimated mean mu hat. The estimated variance is proportional to the estimated mean.

Figure 12. Segment 32 Poisson GLM Estimated Variance vs. Mean Plot.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX E. TOP FIVE PRIZM NE MARKET SEGMENTS AND SEGMENT 47 MODEL RESULTS

Tables 15 and 16 highlight the variables selected for the top five Poisson GLMs and the top five gradient boosted decision tree models respectively.

Table 15. Top Market Segments and Segment 47 GLM Model Performance.

PRIZM NE Segment	Predictor Variables	Pseudo R-Squared
Segment 32	non_hs_grad_18to24, somecollege_18to24, unemployed, armed_forces_employed, arts, public_admin, commute30to60, non_fam_alone, vacant_units, Veterans, median_income, Num_Schools, Non-violent, Total_Deaths, Stroke, No_Exercise, Few_Fruit_Veg, Smoker, Diabetes	0.521
Segment 20	Enrolled_college_18to24, armed_forces_employed, public_admin, non_fam_roomate, Non_violent, Lung_Cancer, Unhealthy_Days, Toxic_Chem	0.219
Segment 33	Enrolled_college_18to24, armed_forces_employed, aggriculture, arts, percapita_income, Democrat, Non_Violent, Lung_Cancer, Total_Deaths, No_Exercise, Smoker, Uninsured, Toxic_Chem	0.38
Segment 37	Enrolled_college_18to24, aggriculture, manufacture, median_income, Democrat,	0.313

	Lung_Cancer, Total_Deaths, Smoker, Uninsured, Unhealthy_Days, Toxic_Chem	
Segment 47	armed_forces_employed, commute_60plus, Veterans, median_income, Lung_Cancer, Total_Deaths, Few_Fruit_Veg, Obesity, Recent_Drug_Use	0.3578
Segment 63	armed_forces_employed, percapita_income, Num_Schools, Uninsured	0.351

Table 16. Top Market Segments and Segment 47 Gradient Boosted Decision Tree Model Performance.

	Predictor Variables	**Influence**	**Direction**
Segment 20	armed_forces_employed	7.625	Positive
	Total_Deaths	5.864	Positive
	Lung_Cancer	5.685	Positive
	Total_Births	4.152	Positive
	median_income	3.896	Positive
	Stroke	3.798	Negative
	Toxic_Chem	3.291	Neutral
	High_Blood_Pres	3.230	Positive
	Diabetes	2.979	Negative
	Homicide	2.843	Positive
	Predictor Variables	**Influence**	**Direction**
Segment 33	median_income	7.749	Negative
	aggriculture	7.439	Negative
	Lung_Cancer	6.839	Positive
	Smoker	5.088	Positive
	percapita_income	5.038	Negative
	Few_Fruit_Veg	4.801	Negative
	Suicide	3.803	Neutral

	Predictor Variables	Influence	Direction
	Total_Births	3.739	Positive
	Total_Deaths	3.500	Negative
	Uninsured	2.486	Negative

	Predictor Variables	Influence	Direction
Segment 37	Total Deaths	11.453	Negative
	aggriculture	5.196	Negative
	armed_forces_employed	4.639	Positive
	Total Births	4.097	Negative
	HH_income_200000plus	3.981	Negative
	below_poverty	3.424	Negative
	Lung Cancer	3.213	Positive
	percapita_income	3.115	Negative
	housing_units	2.961	Positive
	Uninsured	2.797	Negative

	Predictor Variables	Influence	Direction
Segment 47	High_Blood_Pres	6.287	Negative
	aggriculture	5.855	Negative
	Major_Depression	5.689	Negative
	Manufacture	5.125	Negative
	Recent_Drug_Use	4.298	Positive
	Somecollege_18to24	3.853	Positive
	Commute_60plus	3.291	Negative
	percapita_income	2.970	Negative
	Smoker	2.845	Neutral
	Uninsured	2.656	Positive

	Predictor Variables	Influence	Direction
Segment 63	Uninsured	21.998	Positive
	armed_forces_employed	17.794	Positive
	College_grad_18to24	8.568	Negative
	Major_Depression	5.592	Negative
	Lung Cancer	3.857	Negative
	No_Exercise	3.801	Positive
	Total Deaths	2.973	Negative
	Suicide	2.806	Positive
	Enrolled_college_18to24	2.592	Negative
	vacant_units	2.116	Negative

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Centers for Disease Control and Prevention. (2010). Community health status indicators to combat obesity, Heart Disease and Cancer. Retrieved from http://www.healthdata.gov/dataset/community-health-status-indicators-chsi-combat-obesity-heart-disease-and-cancer

CHSI. (2016). Information for improving community health. Retrieved from http://wwwn.cdc.gov/CommunityHealth/home

Clingan, L. (2011). SAMA not just another acronym. *Recruiter Journal, 63*(6), 27.

Faraway, J. (2006). *Extending the linear model with R.* Boca Raton, FL: Taylor and Francis Group.

Feeney, N. (2014, June 29). Pentagon: 7 in10 youths would fail to qualify for military service. *Time.* Retrieved from http://time.com/2938158/youth-fail-to-qualify-military-service/

Friedman, J., Hastie, T. , & Tibshirani R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*(1), 1.

Housing and Urban Development (2016a). County—ZIP crosswalk, 4th quarter 2015. Retrieved from https://www.huduser.gov/portal/datasets/usps_crosswalk.html

Housing and Urban Development (2016b). FBI data. Retrieved from http://socds.huduser.gov/FBI/FBI_Home.htm?

Intrater, B. C. (2015). Understanding the impact of socio-economic factors on Navy accessions (Master's thesis). Naval Postgraduate School. Retrieved from Calhoun http://hdl.handle.net/10945/47279

Jackson, Sandra Y. 2015. "Utilizing socio-economic factors to evaluate recruiting potential for a U.S. Army Recruiting Company." Master's thesis, University of Texas, Austin.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2015). *An introduction to statistical learning with applications in R.* New York: Springer.

Joint Advertising Market Research & Studies (2014). *Your guide to more effective recruiting 2014*. Arlington, VA: U.S. Department of Defense.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B. (2016). Caret: classification and regression training. R package version 6.0-68. https://CRAN.R-project.org/package=caret

Marmion, W. N. (2015). *Evaluating and improving the SAMA (segmentation analysis and market assessment) recruiting model.* Master's thesis, Naval Postgraduate School. Retrieved from Calhoun http://hdl.handle.net/10945/45894

McHugh, John M., & Raymond T. Odierno. *A statement on the posture of the United States Army, fiscal year 2015*. Posture Statement presented to the 114th Cong., 1st sess. Washington, DC: U.S. Department of the Army, 2015.

Metcalf, P. A., Scragg, R. R. K., Schaaf, D., Dyall, L., Black, P. N., & Jackson, R. T. (2008). Comparison of different markers of socioeconomic status with cardiovascular disease and diabetes risk factors in the diabetes, heart and health survey. *The New Zealand Medical Journal*, *121*(1269), 45–56. Retrieved from http://libproxy.nps.edu/login?url=http://search.proquest.com.libproxy.nps.edu/docview/1034236572?accountid=12702

National Center for Education Statistics (2016). *Integrated postsecondary education data system*. Retrieved from http://nces.ed.gov/ipeds/.

Nielsen. (2016, April 09). MyBestSegments. Retrieved from www.nielsen.com: https://segmentationsolutions.nielsen.com/mybestsegments/Default.jsp?ID=70&&pageName=Learn%2BMore&menuOption=learnmore.

Parker, N. (2015). Improved Army Reserve unit stationing using market demographics. Naval Postgraduate School. Retrieved from Calhoun http://calhoun.nps.edu/handle/10945/45921

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Ridgeway, G. (2015). Generalized boosted regression models. R package version 4.1-3. https://cran.r-project.org/web/packages/gbm/gbm.pdf.

Tice, J. (2016, February 23). Army recruiting market tightens but service expects to make 2016 goal. *Army Times.* Retrieved from http://www.armytimes.com/story/military/careers/army/2016/02/23/army-recruiting-market-tightens-but-service-expects-make-2016-goal/80624982/

United States Census Bureau (2013). Maps of metropolitan and micropolitan statistical areas. Retrieved from http://www.census.gov/population/metro/data/maps.html.

United States Census Bureau (2016a). American community survey. Retrieved from http://www.census.gov/programs-surveys/acs/about.html

United States Census Bureau (2016b). Cartographic boundary Shapefiles—ZIP code tabulation areas (ZCTAs). Retrieved from https://www.census.gov/geo/maps-data/data/cbf/cbf_zcta.html

United States Office of Management and Budget (2015). Revised delineations of metropolitan statistical areas, micropolitan statistical areas, and combined statistical areas, and guidance on uses of the delineations of these areas. OMB Bulletin No. 15–01. Retrieved from https://www.whitehouse.gov/sites/default/files/omb/bulletins/2015/15-01.pdf

United States Postal Service Office of Inspector General (2013). *The untold story of the ZIP code*. Retrieved from http://postalmuseum.si.edu/research/pdfs/ZIP_Code_rarc-wp-13-006.pdf

USAREC. (2009). *Recruiting operations*. Fort Knox, KY: United States Army Recruiting Command.

USAREC. (2012). *Recruiting center operations*. Fort Knox: United States Army Recruiting Command.

USAREC. (2015). About us. Retrieved from USAREC: http://www.usarec.army.mil/aboutus.html

USAREC G2. (2012). Segmentation analysis and market assessment (SAMA) reports user guide. Fort Knox: USAREC G-2.

Zhao, Y., Cen, Y. (2014). *Data mining applications with R*. Oxford: Elsevier.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California